Données anonymisées, données pseudonymisées : de quoi s'agit-il ?

Les données sont à caractère personnel dès lors qu'elles concernent des personnes physiques, identifiées directement (nom, prénom) ou indirectement (numéro de sécurité sociale, identifiant national de santé, numéro de téléphone, empreinte digitale, etc.).

On distingue :

- ✓ **l'anonymisation irréversible** : consiste à supprimer tout caractère identifiant aux données personnelles. Toutes les informations directement et indirectement identifiantes sont supprimées de façon à rendre impossible toute ré identification des personnes.
- ✓ **l'anonymisation réversible, ou « pseudonymisation »** : consiste à remplacer un identifiant par un pseudonyme. Ceci permet la levée de l'anonymat ou l'étude de corrélations en cas de besoin. Les données pseudonymisées ne sont pas considérées comme anonymes.

Pourquoi faut-il anonymiser les données personnelles ?

Dès qu'une entreprise traite des données personnelles, **des règles de confidentialité et de sécurité très contraignantes sont imposées par la loi Informatique et libertés**, notamment afin d'empêcher que des tiers non autorisés accèdent aux données personnelles.

Ces règles sont d'autant plus contraignantes lorsqu'il s'agit de **données de santé**, qualifiées de sensibles par la loi IFL.

Le phénomène des « **données massives** » (quantité exponentielle des volumes de données personnelles numériques collectées et traitées) met en risque la sécurité des données personnelles.

- ➔ Les techniques d'anonymisation peuvent répondre à ces règles de sécurité, et apporter des **garanties en matière de respect de la vie privée** des personnes
- ➔ **Une anonymisation irréversible des données personnelles permet de sortir du périmètre de la loi Informatique Fichiers et Libertés**

Les avantages offerts par l'ouverture des données sont multiples (ex : open data, statistiques) à condition que soit respecté le droit à la protection de ses données personnelles et à la vie privée.

L'objectif de l'anonymisation : protéger la vie privée tout en rendant les données accessibles aux activités qui peuvent en nécessiter, telles que la recherche en matière de santé publique.

Les solutions d'anonymisation sont donc aujourd'hui essentielles pour de nombreuses entreprises qui souhaitent valoriser leurs informations.

Comment s'assurer de l'efficacité de la solution d'anonymisation mise en place ?

En amont, les données personnelles doivent avoir été collectées et traitées dans le respect de la législation applicable.

Trois critères ont été établis par le G29 (Groupe regroupant les autorités de protection des données européennes) pour déterminer si une solution d'anonymisation est efficace et répond aux contraintes légales :

- **L'individualisation** : est-il toujours possible d'isoler un individu ?
- **La corrélation** : est-il possible de relier entre eux des ensembles de données distincts concernant un même individu ?
- **L'inférence** : peut-on déduire de l'information sur un individu ?

Un ensemble de données pour lequel au moins un des trois critères n'est pas respecté ne pourra être considéré comme définitivement anonyme qu'à la suite d'une **analyse détaillée des risques de ré-identification**

Tout responsable de traitement doit effectuer une **veille régulière** de ses solutions d'anonymisation pour préserver dans le temps le caractère anonyme des données.

Compte tenu de l'évolution des technologies, des risques résiduels de ré-identification peuvent toujours survenir. Le G29 affirme à ce titre **qu'aucune technique n'est, en soi, infaillible**.

Quels sont les procédés pour anonymiser les données?

Plusieurs techniques d'anonymisation peuvent être envisagées, avec des degrés de fiabilité variables.

Une technique permet rarement à elle seule d'anonymiser totalement des données ; il convient de **combiner plusieurs techniques** (au moins deux).

Il faut tenir compte du facteur de risque et **évaluer la gravité et la probabilité de ce risque pour apprécier la validité d'une technique d'anonymisation (risque = probabilité x impact)**.

En synthèse, les techniques proposées sont les suivantes :

- **Randomisation** : techniques qui altèrent la véracité des données dans le but de supprimer le lien fort entre les données et la personne (ex. : Confidentialité différentielle, Permutation)
- **Généralisation** : technique qui consiste à généraliser ou diluer les données personnelles de façon à ce qu'elles perdent en précision et qu'elles ne soient plus spécifiques à une personne mais communes à un ensemble (ex. : Agrégation et k-anonymat ; l-diversité/t-proximité)